

Impact von Open-Access-Publikationen

Sabine Henneberger | shenneberger@cms.hu-berlin.de

Der Impact wissenschaftlicher Publikationen

Im Zusammenhang mit wissenschaftlichen Publikationen spricht man von Impact, wenn man die, vor allem kurzfristige, Wirkung in Fachkreisen meint. Diese, so die Annahme, wird widergespiegelt, wenn sich Autor(inn)en auf vorangegangene Publikationen beziehen, indem sie diese zitieren. Um ein Maß für den Impact zu bilden, wird zumeist die Anzahl der Zitationen in nachfolgenden Publikationen hinzugezogen. Ein Grund dafür ist, dass sich Zitationen einfach durch Zählung messen lassen. Die Bedeutung, die auf Zitationen beruhende Impact-Maße als Wissenschaftsindikatoren bei der Beurteilung von Publikationen und damit der Leistungsfähigkeit von Autorinnen und Autoren bis heute erlangt haben, hängt mit dieser einfachen Möglichkeit, die durch elektronische Datenverarbeitung auf Datenbanken gestützt automatisch durchgeführt wird, zusammen, andererseits aber auch mit der Annahme, dass mit Zitationen ein Belohnungssystem in der Wissenschaft realisiert wird [1].

Ausführlicheres über Ideen, Hintergründe und unterschiedliche Ansätze der Zitationsanalyse findet man im vorangegangenen Artikel dieses Journals.

Obwohl auf Zitationen beruhende Impact-Maße (Citation Impact) umstritten und ihre Verwendung als ausschließliches Qualitätsmerkmal äußerst fragwürdig ist, spielen sie doch eine herausragende Rolle bei der Bewertung wissenschaftlicher Leistung. Dem liegt die Idee zugrunde, dass Zahlen, gewonnen aus bibliometrischen Analysen, also Statistiken von

Zitationen, objektiv und deshalb besser geeignet sind als z. B. die Einschätzung von Forschungsleistungen durch Peer Review. Zusätzlich dazu, dass Zahlen nur einen Teilaspekt wissenschaftlicher Leistungen abbilden können, werden Statistiken einerseits häufig falsch angewendet und andererseits Ergebnisse falsch interpretiert. Die Folge davon ist, dass die vermeintliche Objektivität der Zahlen oft nur eine Illusion ist. Zu dieser Aussage gelangen die Autorinnen und Autoren des IMU¹-Reports, in welchem eine umfassende kritische Auseinandersetzung mit der Zitationsanalyse geführt wird [2].

Eine Alternative ist die Verwendung von Nutzungsdaten von Publikationen zur Bewertung des Impacts. Nutzungsdaten können von traditionellen, also ausschließlich gedruckten Publikationen erhoben werden, wesentlich besser aber von digitalen Publikationen, die hier im Mittelpunkt stehen.

Die Nutzung digitaler Quellen, die durch elektronische Systeme protokolliert wird, kann auf der Basis von Netzwerkmodellen, die aus den Sozialwissenschaften bekannt sind, analysiert werden. Beispiele sind der PageRank-Algorithmus (Google²) oder die Analyse von Nutzungsdaten in Form von „Usage Networks“ oder „Reader Generated Networks“ [3,4]. Es gibt eine Reihe von Vorschlägen, wie auf der Basis der Analyse von Nutzungsdaten Maße für den Impact von Publikationen definiert werden können.

Solche Alternativen zum Citation Impact sind die von Bollen et al. vorgeschlagenen Metriken für Journale [5].

Impact wissenschaftlicher Publikationen bedeutet nicht nur Citation Impact. Auch der Download Impact kann als ein Maß dafür angesehen werden, wie eine Publikation in Fachkreisen Verbreitung findet. Die Bestimmung des Download Impact aus Protokollen elektronischer Systeme, die Open-Access-Publikationen über das Internet anbieten, ist jedoch nicht unproblematisch und muss normiert werden. Das DFG-Projekt „OA-Statistik“ hat sich dieser Aufgabe gestellt.

1 IMU: International Mathematical Union

2 Google: <http://www.google.de>

Nutzungsdaten sind unter anderem die Downloadzahlen von digitalen Publikationen. Jeder Zugriff eines Internetnutzers auf eine Website, speziell auf eine Publikation, mittels eines Browsers, ist solch ein Download, egal, ob die Daten nur im Browser sichtbar sind oder gespeichert werden. Die Begriffe Downloadzahl und Zugriffszahl werden synonym verwendet. Die Anzahl der Downloads einer digitalen Publikation wird als Download Impact bezeichnet.

Wer auf dem edoc-Server der Humboldt-Universität publiziert, kann sich „seine“ Downloadzahlen ansehen. Sie werden auf der sogenannten „Frontdoor“, der Eingangsseite zur Publikation, veröffentlicht. Die „Frontdoor“, auch als „Abstract“ bezeichnet, erreicht man z. B. über die Homepage des edoc-Servers³ im Bereich „Lesen“ oder über die Suche.

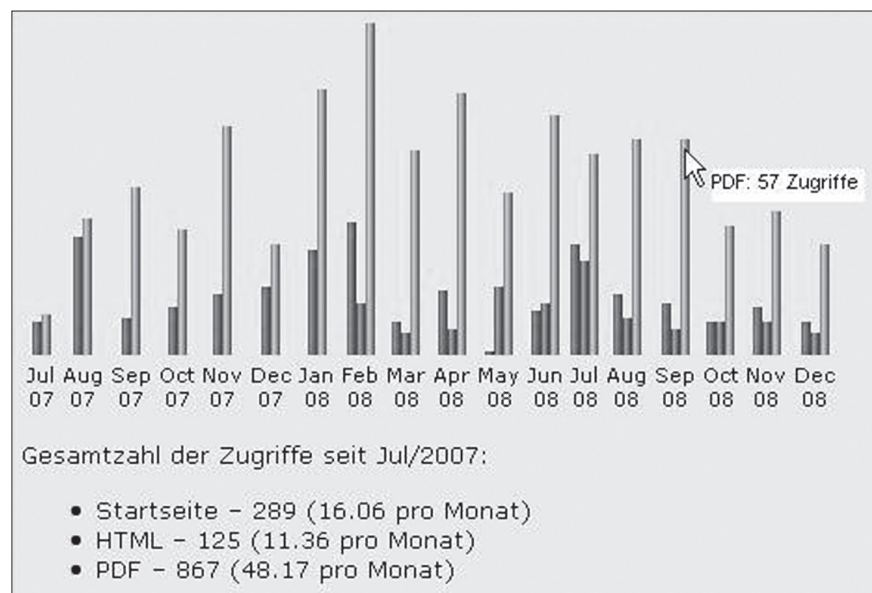


Abb. 1: Zugriffsstatistik auf dem edoc-Server der Humboldt-Universität

Diese Anzeige der Downloadzahlen ist ein Service für die Autorinnen und Autoren, welcher bei weitem nicht von allen Open Access Repositories geboten wird.

Was sagen diese Zahlen aber aus und sind sie wirklich relevant für Autorinnen und Autoren?

Impact und Open Access

Gibt es einen Zusammenhang zwischen Download Impact und Citation Impact? Diese Frage werden Autorinnen und Autoren, besonders wenn sie auf einem Open Access Repository publizieren, stellen, wird ihre wissenschaftliche Leistung doch nach wie vor auf Grundlage von Zitationen in nachfolgenden Publikationen gemessen.

Brody et al. wiesen die Korrelation von Download Impact (Zugriffshäufigkeit), gemessen ab dem Zeitpunkt der Veröffentlichung auf arXiv.org⁴, einem Open Access Repository für Physik, Mathematik, Informatik, Biologie und Statistik, und späterem Citation Impact (Zitationshäufigkeit) von Artikeln auf einigen Gebieten der Physik und Mathematik nach und folgern daraus, dass der

bald vom „Open Access Advantage“ im Hinblick auf den Citation Impact, zumindest für Artikel, die Rede.

Allerdings kann diese Aussage nicht in der Allgemeinheit, wie sie oft verwendet wird, aufrecht erhalten werden.

So wurde der Einfluss von Open Access auf den Citation Impact von Artikeln eines astrophysikalischen Journals untersucht. Die Autoren stellten fest, dass die Open-Access-Publikation keinen höheren Citation Impact zur Folge hatte. Sie kommen zu dem Schluss, dass es viele Argumente für Open Access gibt, von einem generellen „Open Access Citation Advantage“ aber nicht die Rede sein kann [7].

Ebenso beobachteten Davis et al. zwar wesentlich größere Downloadzahlen für Open-Access-Artikel, nicht aber höhere Zitationszahlen [8].

Diese letzte Arbeit führte zu heftigen Kontroversen⁵, nicht zuletzt in methodischer Hinsicht, die einen Einblick in die oft auch polemisch geführte Diskussion zum Thema gibt.

Eine gute Möglichkeit, die Entwicklung auf diesem Gebiet zu verfolgen, bietet die Website des Open Citation Project.⁶

Es scheint im Moment nicht sicher, ob die Veröffentlichung auf einem Open Access Repository dazu führt, dass eine Publikation häufiger zitiert wird, als würde sie nur Non Open Access veröffentlicht werden. Es gibt Beispiele, die dafür, und solche, die dagegen sprechen.

Aber es ist klar, dass Open Access zu höheren Downloadzahlen und damit zu einer besseren Verbreitung wissenschaftlicher Ergebnisse führt. Impact kann eben nicht nur im Sinne von Citation Impact verstanden werden.

⁵ siehe http://www.bmj.com/cgi/eletters/337/jul31_1/a568#199775, gelesen 14.8.2008

⁶ The effect of open access and downloads ('hits') on citation impact: a bibliography of studies, <http://opcit.eprints.org/oacitation-biblio.html>, gelesen 22.12.08

³ <http://edoc.hu-berlin.de>

⁴ <http://archivx.org/>

Das Problem des Download Impacts

Was verbirgt sich genau hinter den Downloadzahlen?

Download- oder Zugriffszahlen einer Publikation werden ermittelt, indem Protokolle, die so genannten Logfiles des elektronischen Systems, mit welchem die Publikation im Internet angeboten wird, ausgewertet werden.

Dafür existiert eine Vielzahl von Programmen, die Downloadstatistiken von Repositories durchführen und eine Gemeinsamkeit besitzen:

Die Downloadzahl einer Publikation ergibt sich aus der Anzahl der Datensätze des Logfiles, die einen Zugriff auf diese Publikation repräsentieren und nicht durch Algorithmus oder Konfiguration ausgeschlossen werden.

Diese Algorithmen differieren erheblich, so dass nicht davon ausgegangen werden kann, dass Ergebnisse unterschiedlicher Programme vergleichbar sind. Selbst bei der Verwendung des gleichen Analyseprogramms für verschiedene Repositories sind durch die vielfältigen Konfigurationsmöglichkeiten keine verlässlichen Aussagen darüber zu treffen, ob die Downloadhäufigkeiten auf die Dokumente des einen größer als auf die des anderen Repository sind.

Von besonderer Bedeutung ist die Filterung maschineller Zugriffe. Solche Robot-Zugriffe machen oft mehr als 50% alle Zugriffe aus und sollen natürlich möglichst ausgeschlossen werden. Dafür gibt es eine Reihe von Prinzipien, nach denen vorgegangen wird. Außerdem existieren Listen bekannter Robots, anhand derer ein Robot ausgeschlossen werden kann, wenn er sich zu erkennen gibt. Sowohl die Verwendung von Robotlisten als auch die Befolgung von Prinzipien zur Robotererkennung können keinen vollständigen Ausschluss maschineller Zugriffe garantieren. Es kann immer nur eine möglichst gute Annäherung erreicht werden.

Solche Downloadzahlen, die stark davon abhängen, unter welchen Bedingungen sie ermittelt wurden, taugen kaum als Maß eines Download Impacts.

Ein Versuch, Downloadzahlen vergleichbar zu machen, wird im Projekt

COUNTER⁷ unternommen. Dieses Projekt startete im Jahr 2002 mit dem Ziel, den damaligen Hauptakteuren unter den Anbietern von Online-Ressourcen, Herausgebern und Bibliothekaren, verlässliche Zahlen über die Nutzung ihrer Produkte und Bibliotheksbestände zu bieten. Die dort aufgestellten Richtlinien bilden inzwischen einen gewissen Standard, reichen aber für die Normierung der Downloadstatistik für Open Access Repositories nicht aus, da zum Beispiel die Filterung von Robotzugriffen hier keine Rolle spielt. Weitere Normierungsbestrebungen für Downloadstatistiken werden z. B. im Projekt LogEc⁸ und durch die IFBAC⁹ unternommen.

Der aktuelle Situation für Open Access Repositories ist so, dass die Downloadzahlen von Publikationen innerhalb eines Repository verglichen werden können, da sie unter identischen Bedingungen ermittelt wurden und sich ein Download Impact immer nur auf genau dieses eine Repository bezieht.

Ein Maß für den Download Impact einer Publikation kann so aber nicht bestimmt werden. Es muss als Mindestanforderung möglich sein, Downloadzahlen so zu ermitteln, dass sie als Download Impact einer Gruppe von Repositories angesehen werden können.

Für Autorinnen und Autoren bedeutet dies im Moment, dass die angegebenen Downloadzahlen Repository-abhängig sind. So können auch die durchschnittlichen Downloadzahlen eines anderen Repository höher oder niedriger ausfallen. Sie können aber mit den Downloadzahlen anderer Publikationen des gleichen Repository verglichen und als Hinweis auf die Verbreitung der Publikation dienen.

Eine Lösung

Downloadstatistiken von Open Access Repositories vergleichbar zu machen, ist das Anliegen des DFG-Projektes "Infra-

struktur für standardisierte Nutzungsstatistiken unter besonderer Berücksichtigung Institutioneller Repositorien"¹⁰, kurz OA-Statistik. In diesem Projekt arbeiten 4 Institutionen zusammen, die DINI-zertifizierte Repositories¹¹ mit unterschiedlicher Software betreiben. Das sind DSpace, OPUS und edoc, die Software des edoc-Servers der Humboldt-Universität. Weiterhin kommen Linkresolver, die bibliographische Daten mit Volltexten verknüpfen (z. B. SFX¹²), und Lizenzserver zum Einsatz.

Um den international aktuellen Stand bei der Ermittlung von Downloadstatistiken zu diskutieren, wurde an der Humboldt-Universität ein Workshop¹³ durchgeführt, an welchem viele namhafte Vertreter von Projekten und Institutionen, die sich mit dem Thema beschäftigen, teilnahmen.

Im Projekt „OA-Statistik“ wird folgendermaßen vorgegangen: Die entstehenden unterschiedlichen Protokolle, die Logfiles, werden lokal, also in der beteiligten Institution, sowohl standardisiert als auch individuell aufbereitet. Standardisiert wird das Verfahren zur Unterscheidung der „Usage Events“, individuell sind die Verfahren zur Zuordnung der „Usage Events“ zu Publikationen, was in den unterschiedlichen Datenstrukturen der Repositories begründet ist. Für die Aufbereitung wird unterschiedliche Software verwendet, im Moment ist der Einsatz von Perl und PHP geplant. Die so aufbereiteten Usage Events werden in Form von OpenURL ContextObjects¹⁴ einem zentralen Aggregator zur Verfügung gestellt. Dieser Aggregator übernimmt nun die eigentliche Arbeit, die Ermittlung der Downloadzahlen für die Publikationen aller 4 beteiligten Institutionen unter gleichen Bedingungen. Hierbei sollen die bereits vorhandenen Standards weitestgehend angewendet werden. Die so bestimmten Download-

¹⁰ <http://www.dini.de/projekte/oa-statistik/>, gelesen 21.01.09

¹¹ DINI: Deutsche Initiative für Netzwerkinformation e.V., <http://www.dini.de>, gelesen 21.01.09

¹² SFX: <http://www.exlibrisgroup.com/sfx.htm>, gelesen 17.2.2008

¹³ <http://www.dini.de/veranstaltungen/workshops/jisc-workshop-2008/>, gelesen 22.01.2009

¹⁴ <http://www.exlibrisgroup.com/category/sfxopenurlsyntax>, gelesen 22.01.2009

⁷ COUNTER: Counting Online Usage Of Networked Electronic Resources, <http://www.projectcounter.org>, gelesen 17.2.2008

⁸ LogEc: <http://logec.repec.org/>, gelesen 28.12.08

⁹ IFABC: International Federation of Audit Bureau of Circulations <http://www.ifabc.org/standards.htm>, gelesen 22.12.08

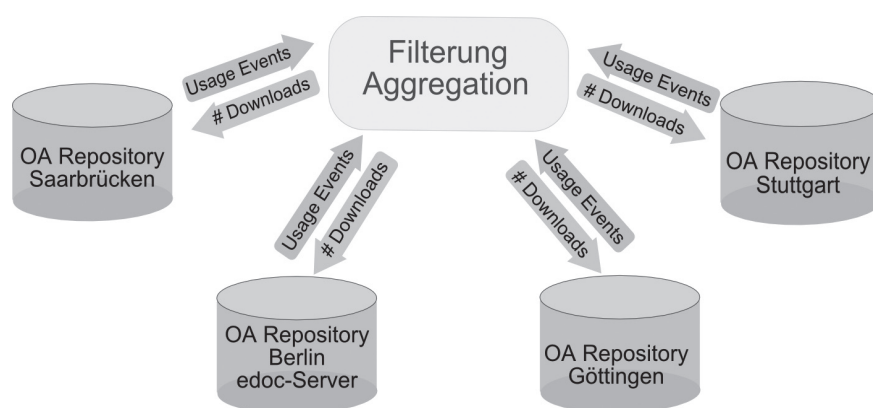


Abb. 2: Überblick über das DFG-Projekt „OA-Statistik“

zahlen der Publikationen werden an die einzelnen Repository-Betreiber zurückgeliefert und in den Metadatenbestand aufgenommen. Von dort aus werden sie über das Internet als zusätzlicher Service für die Autorinnen und Autoren angeboten.

Aber nicht nur die technische Realisierung ist Gegenstand des Projektes. In einem eigenen Arbeitspaket werden Fragen des Datenschutzes geklärt. Das Projekt OA-Statistik kooperiert außerdem eng mit dem DFG-Projekt „Aufbau eines Netzwerks zertifizierter Open Access-Repositories“¹⁵ und liefert mit dem Anbieten von Downloadzahlen einen wichtigen Mehrwertdienst.

Die Ergebnisse des Projektes, mit denen im Frühjahr 2010 zu rechnen sein wird, sollen einerseits in Form von Empfehlungen in das DINI-Zertifikat aufgenommen werden, um eine möglichst weite Verbreitung zu erreichen, andererseits soll die entstandene Software einschließlich Anleitungen zu ihrem Gebrauch Betreibern weiterer Repositories zur Verfügung gestellt werden, die dann ebenfalls in die Lage versetzt werden können, ihren Autorinnen und Autoren vergleichbare Downloadzahlen anzubieten.

Literatur

- [1] ADLER, ROBERT; EWING, JOHN UND TAYLOR, PETER: *Citation Statistics*. A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS), 2008, <http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf>, gelesen am 28.12.08
- [2] BOLLEN, JOHAN; LUCE, RICK; VEMULAPALLI, SOMA SEKHARA UND XU, WEINING: *Usage Analysis for the Identification of Research Trends in Digital Libraries*. D-Lib Magazine Volume 9, Number 5, 2003, [ISSN 1082-9873]; <http://www.dlib.org/dlib/may03/bollen/05bollen.html>, gelesen am 22.01.2009
- [3] BOLLEN, JOHANN UND VAN DE SOMPEL, HERBERT: *Mapping the structure of science through usage*, *Scientometrics* 69, No. 2, 2006, S. 227–258.
- [4] BOLLEN, JOHANN; VAN DE SOMPEL, HERBERT; SMITH, JOAN UND LUCE, RICK: *Toward alternative metrics of journal impact: A comparison of download and citation data*. 2005, <http://arxiv.org/abs/cs.DL/0503007>, gelesen am 4.7.08
- [5] BRODY, T.; HARNAD, S. UND CARR, L.: *Earlier Web Usage Statistics as Predictors of Later Citation Impact*, *Journal of the American Association for Information Science and Technology*. (JASIST) 57(8) 57(8), 2006, S. 1060–1072.
- [6] DAVIS; LEWENSTEIN; SIMON; BOOTH UND CONOLLY: *Open access publishing, article downloads, and citations: randomised controlled trial*. 2008, http://www.bmj.com/cgi/content/abstract/337/jul31_1/a568?maxtoshow=&HITS=10&hits=10&RESULTFORMAT=&fulltext=citation&searchid=1&FIRSTINDEX=0&volume=337&resourcetype=HWCIT, gelesen am 13.8.2008
- [7] HORNBOSTEL, STEFAN: *Wissenschaftsindikatoren: Bewertungen in der Wissenschaft*. Westdeutscher Verlag, Opladen, 1997, ISBN: 3-531-12908-2
- [8] KURTZ, MICHAEL J. UND HENNEKEN, EDWIN A.: *Open Access does not increase citations for research articles from The Astrophysical Journal*. 2007, <http://arxiv.org/abs/0709.0896>, gelesen am 22.01.2009

¹⁵ <http://www.dini.de/projekte/oa-netzwerk/>